📊 Phase 2: Data Collection, EDA, and Hypothesis Testing

📊 Analysis of Crime Rates and Socioeconomic Factors

🚀 Project Overview

This project investigates the relationship between crime rates and socioeconomic factors such as education levels and unemployment rates across various U.S. cities and states. The goal is to identify potential correlations that could inform policies aimed at reducing crime.

🗁 Data Collection

- **Crime Data Source:**

FBI Crime Data Explorer (2023) — City-level crime statistics for violent and property crimes.

- **Socioeconomic Data Sources:**

USDA Economic Research Service (2023) — State-level data on:

- % of adults with a bachelor's degree or higher.
- Unemployment rates.
- The raw crime data from the FBI was cleaned, including removal of redundant rows, handling missing values, and standardizing formats.

Socioeconomic data was filtered and aggregated to match the crime dataset structure.

The two datasets were merged based on state and city to enhance the analysis.

The final processed datasets are available in the /data folder.

📊Exploratory Data Analysis (EDA)

To understand the relationship between crime rates and socioeconomic factors, several exploratory data analysis techniques were applied to the cleaned and merged dataset.

📈Descriptive Statistics:

- The average education level across states was approximately 26.6% holding a bachelor's degree or higher.
- The average unemployment rate stood at 3.8%.

- Crime counts (both violent and property crimes) showed high variability, reflecting differences in population sizes, urbanization, and regional factors.

🔥**Correlation Heatmap:**
A correlation heatmap was generated to visualize the strength and direction of relationships between variables:

- **Education vs. Crime Rates:**
  The heatmap revealed a very weak negative correlation between the percentage of adults with a bachelor's degree and both violent and property crime rates. (**This suggests that higher education levels might slightly align with lower crime, but the relationship is too weak to be considered impactful.)**
- **Unemployment vs. Property Crime:**
  A slight positive correlation was observed, indicating that states with higher unemployment rates tend to experience marginally higher property crime rates.

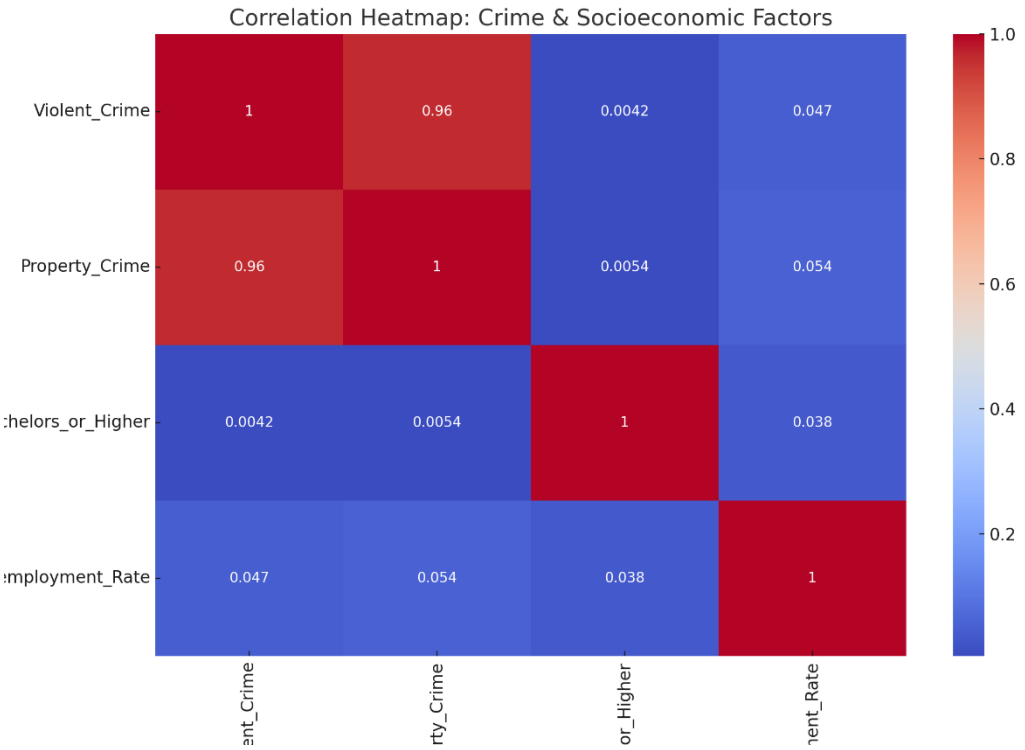
📊**Scatter Plots:**

1. **Violent Crime vs. Education Level**
   This scatter plot showed a wide dispersion of points with no clear trend, confirming that there is no strong linear relationship between education levels and violent crime rates. Even areas with higher education percentages still reported varying violent crime figures, indicating that other factors likely play a larger role.
2. **Property Crime vs. Unemployment Rate**
   The scatter plot displayed a mild upward trend, supporting the correlation result that higher unemployment may be linked to increased property crime. However, the spread of data points still reflects a weak relationship, meaning unemployment alone does not explain property crime rates effectively.

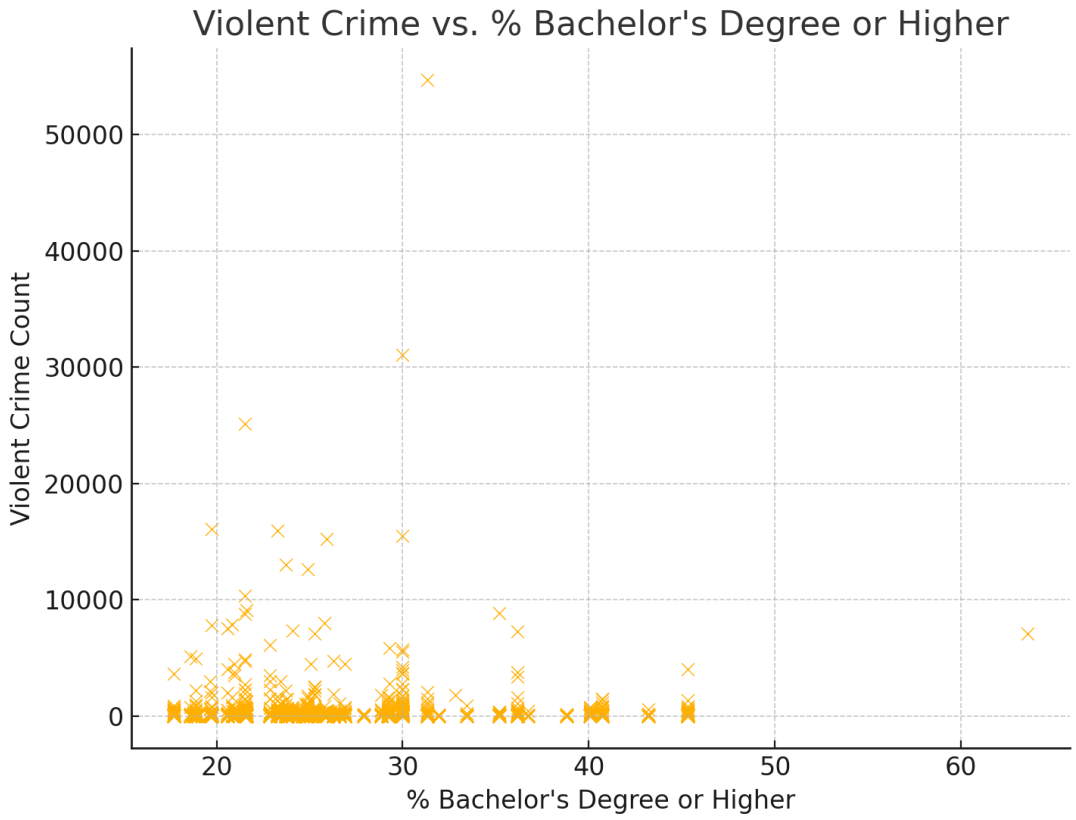
🗁**Visualizations Included:**

- Correlation Heatmap: Visual summary of relationships between variables.
- Violent Crime vs. Education: Scatter plot illustrating lack of clear association.
- Property Crime vs. Unemployment: Scatter plot suggesting a weak positive trend.

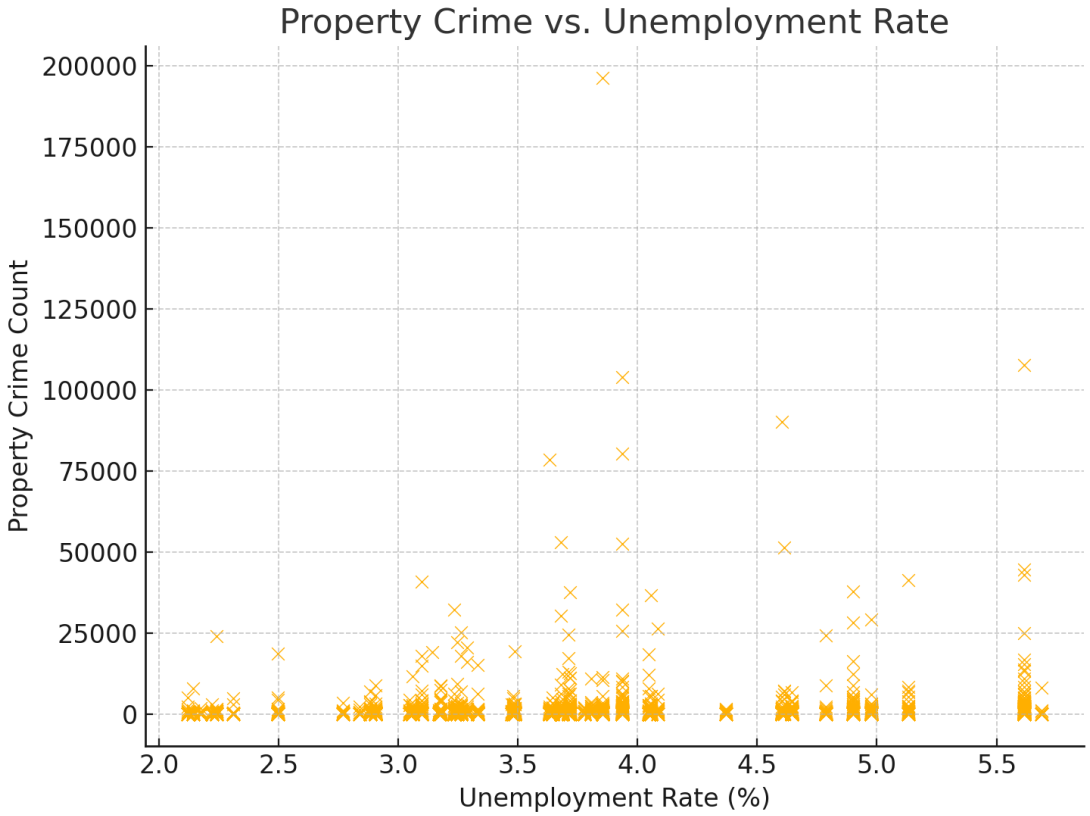
All plots and detailed analyses are available in the "exploratory data analysis" folder.



-



-



More details and plots can be found in the /exploratory data analysis folder.

□ Hypothesis Testing

Two hypotheses were tested:

1□Education vs. Violent Crime

H□: There is no correlation between education level and violent crime.

- **Result:**

Pearson r = 0.0041 (p = 0.706) ➜ No significant linear correlation.

- Spearman ρ = 0.0406 (p < 0.001) ➜ Statistically significant but negligible.

2□Unemployment vs. Property Crime

H□: There is no correlation between unemployment rate and property crime.

- **Result:**

Pearson r = 0.0537 (p < 0.001) ➜ Weak but significant linear correlation.

- Spearman ρ = 0.1453 ($p < 0.001$) ➜ Weak positive monotonic correlation.

📄 Full results are available in the /hypothesis-testing/hypothesis_testing_results.csv.

🎯 Findings & Insights

- Education levels didn't show a strong correlation with violent crime rates in this dataset.
- A weak but statistically significant relationship exists between unemployment rates and property crime.
- These results suggest that while socioeconomic factors may influence crime, they are not sole predictors.

⚠️ Limitations

The crime data used raw counts instead of per capita rates.

- Socioeconomic data was aggregated at the state level, which may overlook local variations.
- Only 2023 data was analyzed; trends over time were not considered.
